American Journal of Medical Genetics (Neuropsychiatric Genetics) 67:254–263 (1996)

An Approach to Investigating Linkage for Bipolar Disorder Using Large Costa Rican Pedigrees

N.B. Freimer, V.I. Reus, M. Escamilla, M. Spesny, L. Smith, S. Service, A. Gallegos, L. Meza, S. Batki, S. Vinogradov, P. Leon, and L.A. Sandkuijl

Center for Neurobiology and Psychiatry (N.B.F., V.I.R., S.V.) and Neurogenetics Laboratory (N.B.F., M.E., L.S., S.S.), Department of Psychiatry (S.B.), University of California, San Francisco, San Francisco, California; University of Costa Rica (M.S., A.G., P.L.), Calderon Guardia Hospital (L.M.), San Jose, Costa Rica; Department of Human Genetics, University of Leiden, The Netherlands (L.A.S.)

Despite the evidence that major gene effects exist for bipolar disorder (BP), efforts to map BP loci have so far been unsuccessful. A strategy for mapping BP loci is described, focused on investigation of large pedigrees from a genetically homogenous population, that of Costa Rica. This approach is based on the use of a conservative definition of the BP phenotype in preparation for whole genome screening with polymorphic markers. Linkage simulation analyses are utilized to indicate the probability of detecting evidence suggestive of linkage, using these pedigrees. These analyses are performed under a series of single locus models, ranging from recessive to nearly dominant, utilizing both lod score and affected pedigree member analyses. Additional calculations demonstrate that with any of the models employed, most of the information for linkage derives from affected rather than unaffected individuals. © 1996 Wiley-Liss, Inc.

KEY WORDS: linkage analysis, isolated populations, bipolar disorder type I, linkage simulation

INTRODUCTION

The availability of detailed genetic maps covering the human genome has led to localization of genes responsible for several complex phenotypes, including neuropsychiatric syndromes such as familial Alzheimer disease. In this paper, we review the obstacles to detecting linkage for bipolar disorder (BP) and suggest strategies for circumventing them. We present the rationale for investigating a population whose history is well documented, describe a set of large families from this population that will be valuable for genetic studies of BP, and discuss the use of computer simulation methods to guide genome screening efforts and linkage analyses.

Mapping genes for common diseases, such as BP, may be complicated by the typically imprecise definition of phenotypes, by etiologic heterogeneity, and by uncertainty about the mode of genetic transmission of the disease trait. However, recent theoretical and empirical findings suggest that even if multiple genes play a role in disease susceptibility, the genes that produce the largest effects can be mapped using procedures that have been applied to single gene disorders [Terwilliger and Ott, 1994; Angrist et al., 1993].

The mapping of a locus for familial malignant melanoma has highlighted the importance of adopting conservative definitions of disease phenotype for linkage studies of common diseases. This gene was mapped by identifying families with several members showing an extreme phenotype, malignant melanoma [Cannon-Albright et al., 1992]. Individuals with a more common but less severe phenotype, cutaneous dysplastic nevi, were considered to have an unknown rather than an affected phenotype; it is now clear that many of these individuals do not carry the disease genotype, and if they had been included as "affected," the linkage findings would have been substantially weaker; a similar finding was obtained in an independent data set [Gruis et al., 1993].

With psychiatric disorders there is even greater ambiguity in distinguishing between individuals who likely carry an affected genotype from those who are genetically normal. For example, one can define an affected phenotype for BP by including one or more of the broad grouping of diagnostic classifications that constitute the mood disorders: BP-I, consisting of at least one episode of full mania, and schizoaffective disorder, manic subtype (SAD-M), are usually considered the most severe syndromes and are the least common in the population. Although major depressive disorder (MDD) and BP-II (MDD with at least one episode of hypomania) are each considerably more common than BP-I and

Received for publication December 13, 1994; revision received October 6, 1995.

Address reprint requests to N.B. Freimer, Neurogenetics Laboratory, UCSF, Box F-0984, San Francisco, CA 94143.

are somewhat less reliably diagnosed [reviewed in Goodwin and Jamison, 1990], individuals with these diagnoses are considered affected by many groups performing linkage studies of BP. Twin studies and recent segregation analyses support the hypothesis that a major locus is involved in transmission of BP-I, but evidence is considerably weaker for a major gene effect for other mood disorders [Bertelsen et al., 1977; Pauls et al., 1992; Spence et al., 1993]; considering these data, our linkage analyses are focused on only those individuals with diagnoses of BP-I and SAD-M, as described in this paper.

The segregation pattern of BP in most, if not all pedigrees is suggestive of incomplete penetrance, and it is thus impossible, with certainty, to correlate phenotypes and genotypes in well individuals. Delayed onset of illness was responsible for much of the loss in evidence for linkage between BP and markers on chromosomes 11 and X in two widely publicized follow-up studies [Kelsoe et al., 1989; Baron et al., 1993]. Although linkage programs can partially correct for mis-assignments based on incomplete penetrance, a more conservative solution is to assign unaffected individuals an "unknown" phenotype. A large study sample is needed to compensate for the loss of statistical power resulting from such a conservative approach.

In linkage analysis of any common disease there is the risk of locus heterogeneity, that is, unlinked susceptibility mutations each accounting for similar phenotypes. Two alternative strategies have been proposed for linkage analysis of BP, given that interfamilial heterogeneity cannot be ruled out: collection of large numbers of multiplex families or investigation of one or more extended pedigrees. Large pedigrees potentially provide more genetic information than do small ones, but there may be a substantial probability of intrafamilial heterogeneity in extended pedigrees especially if they are drawn from large, mixed populations. The chance of such heterogeneity is presumably lower if pedigrees are drawn, as in our study, from isolated populations, in this case, that of Costa Rica.

Most Costa Ricans are descended from a small number of Spanish and Amerindian founders in the 16th and 17th centuries, with little additional immigration until the end of the 19th century. The homogeneity of the population is further indicated by a high prevalence of particular rare recessive diseases in comparison with other populations [Saborio, 1992]. Large sibships remain common in Costa Rica; for example, as of the 1970's, the average 40-year-old woman had given birth to seven children. Because the population density is high and geographical mobility is low, most members of a given family can be readily contacted and evaluated. In addition, it is possible to document the composition of pedigrees over the past 200–300 years through church and civic records and extensive genealogies maintained by many families. These records enable identification, within a given pedigree, of ancestry from more recent immigrants to Costa Rica. In mapping a gene for inherited deafness Leon et al. [1992] showed the advantages of the Costa Rican population for linkage studies. They investigated a single extended pedigree and were able to reconstruct the pattern of disease transmission over more than 200 years through the use of church and legal records. We have used a similar approach to identify and examine extended BP pedigrees in Costa Rica.

Because these pedigrees consist of several hundred living members, it was necessary to identify subsets that would be most informative for linkage analysis. Heutink et al. [1992] showed that linkage simulation studies are quite useful for this purpose for a simple Mendelian disease. For a syndrome such as BP in which genetic complexity is likely, it is important to evaluate the informativeness of such subsets under a range of genetic models. In this paper, we simulate results under several models of single locus inheritance.

METHODS AND RESULTS

Family Descriptions

To date we have investigated four extended Costa Rican pedigrees (Families CR001, CR002, CR003, and CR004), all identified through probands with a clear diagnosis of BP-I. Each of the probands had been clinically evaluated or treated by one of the investigators (NBF, AG, or LM) and had described an extensive family history of BP. In each family, we attempted to systematically collect information about the pedigree structure (and subsequently about affectedness, as described below) beginning with first degree relatives and then extending to progressively more distant ones. Some branches in each of these families remain unexplored.

The size and structure of family CR001 were determined from a family genealogy dating to about 1810 which includes a total of 970 identified descendants of a founding couple and 365 identified "married in" individuals. Ancestry is traced to known individuals who migrated from Spain in the 17th and 18th centuries. For CR002, a family genealogy details the relationship of almost 6,000 individuals (living and dead) beginning with a single male ancestor and his two wives, all three of whom migrated from Spain in the 17th century; this information was collected over several years by a single member of the family and documented in computer files. In family CR003, we have currently identified over 200 individuals descended from a founding couple in the late 18th century. Family CR004, with over 200 identified living members, has been traced to a founding couple at the beginning of the 18th century. However, CR003 and CR004 share a common ancestor in that the founding female of CR004 and the founding male of CR003 were third cousins. For the founding couples in CR003 and CR004, we have been able to identify the majority of ancestors for two to four generations before the starting points of the "identified" families; all of these individuals are of primarily Spanish descent. Although we are now attempting to investigate clinical histories for individuals who connect these families it is difficult to identify the route of transmission of illness in an era before current concepts of mental illness were developed and before adequate records were available. There is also intermarriage between families CR001 and CR004. In CR001, one of the most

densely affected sibships has a deceased mother who carried a clear diagnosis of BP-I based on review of hospital records and a best estimate consensus diagnosis. In this sibship, however, the father is a member of Family CR004. It is noteworthy that no information about these individuals or the heavy loading for BP in family CR004 emerged from standard family histories taken from his descendants in Family CR001; this is in line with previous documentation of under-reporting of psychopathology using this method [Pauls et al., 1990]. The convergence of these families, which had been ascertained independently of each other (as had families CR003 and CR004), was only discovered by piecing together the genealogies. We have also identified two instances of consanguinity within family CR003, and three instances in CR001. CR002 and CR004 are characterized by extensive consanguinity including many cousin matings over successive generations. The documented connections between and within these families confirm our impression about the limited initial population size of Spanish descended Costa Ricans [Escamilla et al., submitted]. Because extensive consanguinity is characteristic of the general population of Costa Rica, its presence in these pedigrees should not be taken as evidence for a recessive mode of transmission of BP.

Diagnostic Procedures

In each branch of each of the pedigrees that we have chosen to investigate there is an identified proband with a diagnosis of BP-I based on semi-structured interview or extensive medical records. As much as possible, diagnostic evaluations of relatives were conducted in a systematic order; the initial person evaluated in each branch of a family would recruit other family members, beginning with first degree relatives. However as social contacts within a family do not necessarily correlate with the degree of biological relationship between individuals, it was often necessary to pursue particular branches based on the interest shown by one or more especially motivated and cooperative individuals. Most interviews were conducted in the homes of the subjects.

Except for a very small number of individuals who are fluent in English, interviews have been conducted in Spanish by bilingual clinicians. All available consenting family members over 18 years of age have been interviewed using the Schedule for Affective Disorders and Schizophrenia-Lifetime version [Endicott and Spitzer, 1979], by a psychologist or psychiatrist trained in the use of this instrument. To date we have interviewed 301 subjects (98 in CR001, 32 in CR002, 19 in CR003, and 152 in CR004). A subset of individuals (n = 87) were interviewed for a second time, usually within one year of the first interview, for the purpose of providing increased certainty regarding diagnoses for the best estimate diagnostic procedure. These individuals received a second interview because of a diagnosis, derived from the first interview, of major affective disorder (BP-I, SAD-M, BP-II, MDD), schizophrenia, or substance abuse or because they were selected to be part of a random sample of unaffected cases constructed to prevent the second interviewers from pre-judging subjects as "affected." These interviews were conducted using the NIMH Diagnostic Instrument for Genetic Studies (DIGS) [Nurnberger et al., 1994] by a psychiatrist who is expert in the diagnosis of mood disorders.

Because Costa Rica is a small country with a relatively non-mobile population, and because there was early recognition of the need for specialty treatment of severely ill psychiatric patients, it is likely that most individuals with definite mania have been hospitalized within the country, except for a few wealthy individuals who may have sought treatment abroad. All Costa Ricans have access to psychiatric care through the country's social security system, which has been in place for several decades [Gallegos, 1988]. The reconstruction of diagnoses through the evaluation of medical records has been facilitated by the fact that from its founding in 1890 until 1970 the National Psychiatric Hospital in San Jose was the only hospital in Costa Rica which accepted psychiatric inpatients; a central alphabetized record file permits evaluation of records from the founding date of the hospital. A small inpatient unit opened in 1970, at the Calderon Guardia Hospital, and also has an alphabetized record system. Medical records have been reviewed and abstracted by a research psychiatrist based on a semi-structured protocol designed to elicit specific symptoms and their duration. A total of 43 records have been examined for individuals who were not available for interviews (22 in CR001, eight in CR002, five in CR003, and eight in CR004).

Best estimate diagnostic procedures were established based on the method of Leckman et al. [1982]. Three research psychiatrists, at the University of California, San Francisco, who are expert in the diagnosis of mood disorders reviewed the cases, with two raters reviewing each case. The best estimators independently reviewed the clinical materials and completed diagnostic checklists, recording symptoms relevant to RDC and DSM-IIIR criteria [Spitzer et al., 1978; American Psychiatric Association, 1987]. In addition, the best estimators recorded their overall clinical impressions on a separate scale. The best estimators then jointly reviewed the diagnoses, and, if there was agreement, recorded a consensus diagnosis. If the best estimators did not reach consensus, which occurred in 10 instances, they requested additional information which was collected by additional interviews, by obtaining more details from family members, or through further review of medical records. Of the individuals who have been evaluated through the best estimate diagnosis we have identified 31 cases of BP-I, and an additional 3 cases of SAD-M; a complete diagnostic summary is in Table I.

Selection of a "Screening Panel" of Pedigree Members

In order to permit efficient genome screening, it was necessary to identify a subset of individuals from families (CR001 and CR004) who would provide a substantial amount of genetic information according to the following plan: first the genome would be screened with a defined set of mapped highly polymorphic markers,

TABLE I. Diagnostic Summary of Individuals Interviewed From the Costa Rican Families

Total subjects	344
First interview diagnoses	301
Bipolar I	26
Bipolar II	29
Bipolar NOS	13
Major depressive disorder	16
Substance abuse/dependence	28
Second interviews	87
Best estimate diagnoses	77
Bipolar I	31
Schizoaffective, manic	03
Bipolar II	03
Major depressive disorder	09
Substance abuse/dependence	23

then more extensive genotyping would be performed (more markers and a larger study sample) in regions surrounding markers with suggestive lod scores (e.g., >1.0). Our goal for the genome screening studies was thus to identify promising regions for further investigation rather than to unequivocally demonstrate linkage.

For initial genome screening we focused on Families CR001 and CR004, based on a higher density of identified BP-I individuals, more complete genealogies, and greater degree of cooperation than in the other families. From the individuals who had been interviewed, we chose 62 for initial genotyping, including all those with best estimate diagnoses of BP-I or SAD-M as well as individuals who would provide the most information on linkage phase within each family. For linkage analyses we also included an additional 65 individuals in these families who were not available for genotyping; these individuals were assigned diagnoses through the best estimate process if sufficient medical records were available, otherwise they were used only to assist in setting the pedigree structure. The portions of Families CR001 and CR004 that were selected for initial linkage analyses are shown in Figure 1. The suitability of this sample for genome screening was evaluated using linkage simulation analyses, with a range of single locus models of genetic transmission.

Selection of Models and Linkage Simulation Studies

There is no clear consensus about how to design appropriate single locus models for linkage analysis of complex traits. Segregation analyses give the best fitting parameters for a single major locus, assuming systematic ascertainment of cases, but these parameters may be unsatisfactory to describe complex modes of inheritance. Therefore, a set of criteria were established, to be met by any model which would be employed for linkage analyses:

1. It must correspond to a realistic expected population prevalence of the disease.
2. It should absorb apparent recombination events in patients who are affected due to causes other than the gene(s) of interest, that is it should allow for a substantial probability that affected individuals represent phenocopies.
3. It should give a reasonable power to detect disease genes.
4. Most of the linkage information should be provided by affected individuals.

For all linkage analyses the population prevalence of BP was set at 0.015, based on epidemiological data for Costa Rica, obtained through a national survey [Adis, 1992]. The population frequency of phenocopies was arbitrarily set at 0.01 (⅔ of the total cases of BP), based on criterion 2.

The power to detect disease genes was investigated by simulating linkage using five single locus models of transmission of BP, ranging from fully recessive to almost dominant. These models differ mainly in the penetrance for the heterozygote (f_1), which is calculated by the following: $f_1 = f_0 + x^*(f_2 - f_0)$, where x is the degree of dominance, f_0 is the penetrance for individuals not carrying the disease gene, f_1 is the penetrance for one copy of the disease gene, and f_2 is the penetrance for two copies of the disease gene. For the five models, x varies from 0 (recessive model, $f_1 = f_0$), to 0.25, 0.50, 0.75, or 0.90 (almost dominant model, $f_1 \sim f_2$). We arbitrarily set f_2 to be 0.9. The prevalence of the disease was calculated by: $f_0^*(1 - q)^2 + f_1^*2q^*(1 - q) + f_2^*q^2$, where q is the frequency of the disease allele. We adjusted q, f_0, and f_1 so that with each increment of x, as described above, the correct prevalence of BP was maintained. The five models with their respective parameters are shown in Table II.

For each of the two families (CR001 and CR004) we used the SLINK program to generate 100 replicates of marker data for each of the five models. The marker had four equally frequent alleles, for a heterozygosity of 0.75 (similar to the average informativeness of markers available for genotyping). Marker data were simulated, for each model/family, at a recombination value of 0.05, the maximum possible distance between a marker and the BP locus given a map of markers at approximately 10 cM intervals, as would be used in actual linkage analysis.

For each of the 100 replicates in each model/family the MLINK option of the LINKAGE package was used to evaluate lod scores at 50 values of theta (0.0–0.49, in steps of 0.01). The power to detect linkage was first calculated separately for each model/family and then the results for two families were combined (for each model), by randomly selecting (with replacement) one of the 100 replicate pedigrees for family CR001, and randomly selecting one replicate for family CR004 and evaluating the lod score in the resultant combined pedigrees. The lod scores at each of the 50 values of theta were summed, and the maximum noted. This procedure was performed 5000 times.

Using the same simulated data, we also investigated the power of the affected pedigree member (APM) method of linkage analysis [Weeks and Lange, 1988]. This method has become popular as an adjunct to the lod score method in the analysis of diseases with complex patterns of inheritance. Specification of parameters involved in disease transmission are not necessary for this method, which is based on evaluating the degree of allele sharing between affected individuals with

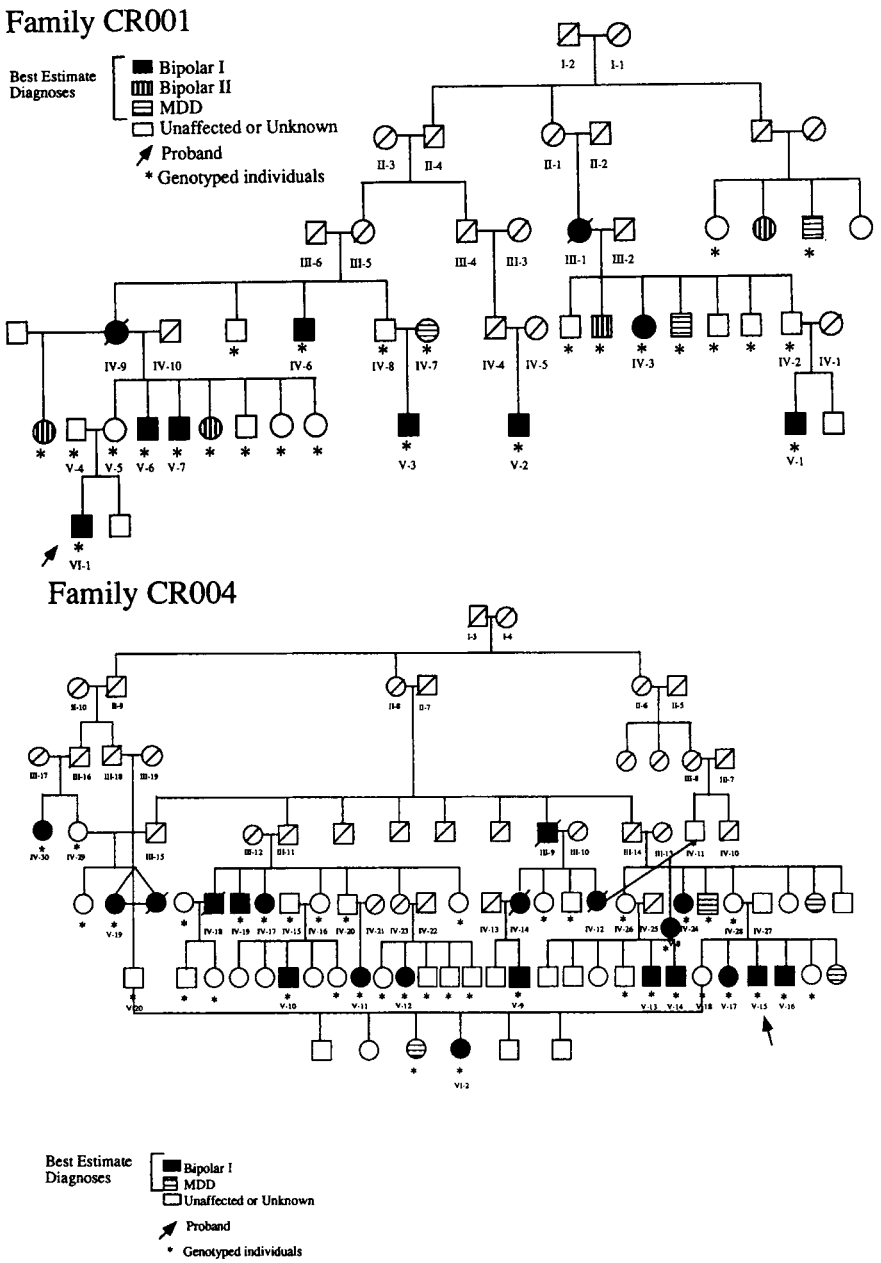

Fig. 1. Pedigrees for Costa Rican families CR001 and CR004 used in the linkage simulation analyses. Probands in each family are indicated by arrows. Roman numerals refer to individuals who have been included in genotyping studies and whose marker haplotypes are depicted in Freimer et al. [1996].

known degree of relationship. Although the degree of similarity is often weighted by marker allele frequency, this was not necessary for these simulations because we considered markers for which all alleles are equally frequent. The significance of allele sharing was evaluated by comparing the test statistic, T, to the quantiles of the standard normal distribution. For each of the five models in both families, 100 T statistics were calculated, and the distribution of these statistics was examined. For each model, we combined families by looking at all 100*100 combinations of the replicates for each family, and calculated a combined T value.

The five models were evaluated to determine, for each of them, the proportion of linkage information that derives from affected and unaffected individuals. Only three mating combinations are informative for linkage with respect to the alleles of the disease gene: 1) homozygous affected with heterozygous, 2) homozygous unaffected with heterozygous, and 3) heterozygous with heterozygous. The frequencies of each of these mating types in the population are found by (1) $4*q^3*(1 - q)$, (2) $4*(1 - q)^3*q$ and (3) $4*q^2*(1 - q)^2$, respectively, where q is the frequency of the disease gene. Furthermore, the probability for a given mating type to

TABLE II. Models of BP Transmission and Their Associated Parameter Values*

Model	x	Disease gene frequency	Penetrance NN (f_0)	Penetrance DN (f_1)	Penetrance DD (f_2)
I	0.00	0.063608	0.011404	0.011404	0.9
II	0.25	0.010639	0.010216	0.232662	0.9
III	0.50	0.005494	0.010110	0.455055	0.9
IV	0.75	0.003694	0.010074	0.677518	0.9
V	0.90	0.003087	0.010062	0.811006	0.9

* In all cases the population prevalence of BP is assumed to be 0.015 and the population frequency of phenocopies to be 0.01. The models refer to the degree of dominance, from fully recessive (I) to nearly dominant (V). Penetrances were established using the relationship $f_1 = f_0 + x*(f_2 - f_0)$ as described in the text.

either produce an affected or an unaffected child was taken into account (e.g., for mating type 2 the probability for an offspring to be affected is $0.5*f_0 + 0.5*f_1$). As these are the only informative mating types, the contribution of a given informative mating type to all informative matings is found by dividing the population frequency of the mating type for a child with a given phenotype by the sum of the population frequencies for all informative matings.

Using these mating type frequencies, analyses were performed between a fully informative marker and a BP locus (that is using the parameters set for models 1–5 as described above). For these analyses a model pedigree was constructed consisting of four grandparents, two parents, and one child; this is the simplest pedigree structure that permits evaluation of the different mating types. Eight different families were created that encompassed all combinations of disease and marker phenotype in the child (two disease phenotypes: affected and unaffected and four possibilities for marker genotype). The lod scores at a recombination fraction of 0.05 were calculated for each of the eight families using the MLINK program of the LINKAGE package. The four lod scores for families with affected children were weighted by the probabilities of occurrence and summed to create an expected lod score (ELOD) for affected children. The same procedure was used on the four lod scores for the unaffected children. This process was repeated for each mating type/model combination to produce 15 affected/unaffected ELODs. A weighted affected and unaffected ELOD score was calculated for each model by multiplying the ELOD by the relative frequencies of each mating type. The ratio of the weighted ELOD for affected children to the weighted ELOD for unaffected children provides an assessment of the amount of information conveyed by an affected child compared to an unaffected child.

The ratio of the information derived from affected versus unaffected individuals was compared to that obtained for analyses utilizing the disease transmission parameters employed in linkage investigations of BP in a large Amish pedigree [Egeland et al., 1987]. In this case, disease prevalence in the population was set at 3.63%; this figure is considerably higher than the one employed for analyses of the Costa Rican BP families because it is based on a less restrictive definition of affectedness. The phenocopy rate was set at 2.6% and the model was completely dominant, with the penetrance for heterozygotes and homozygous affecteds being 0.85.

Linkage Simulation Results

For each replicate (in each model/family) the maximum lod score was noted. On average, the recombination value corresponding to the maximum lod score was near 0.05 for each model/family, as expected. For each model/family, the percentage of lod scores ≥ 3 gives an approximation of the power of a given model to correctly detect significant linkage. The results are shown in Figure 2.

For all models, the expected evidence for linkage is always substantially greater in the sample drawn from Family CR004 than that drawn from CR001, as would be predicted based on the higher number of BP-I individuals in this family (see Fig. 1). For Family CR004 there is a high probability of detecting suggestive lod scores (>1.0) for any of the models, and in some models (e.g., when $x = 0.75$), the power is substantially increased when results for the two families are combined (assuming locus homogeneity). Although the power to detect linkage increases with the degree of dominance, the pattern is not uniform. For example, in Family CR004 the power is greater under the fully recessive model (I) than under the model assuming 25% dominance. Similarly, the power to detect linkage for Family CR001 is about the same between models IV (75% dominance) and V (90% dominant).

In these simulations the power of the APM method to detect linkage is less than for the lod score method, particularly in combining results between families (Table III). It performs almost as well as the lod score method, however, in the more recessive models, especially model II.

Evaluation of the Information From Affected/Unaffected Children

Because of the low frequency of the disease gene, the mating type involving one homozygous unaffected individual and one heterozygous individual (mating type 2) is the most frequent. Since the disease gene frequency under these models decreases with increased dominance, the relative frequency of mating type 2 is highest in the dominant models. ELOD scores, weighted ELOD and the ratio of information provided by affected compared to unaffected individuals are presented in Table IV. Because there is more certainty in the genotype of an unaffected individual with increasing degree of dominance an unaffected child provides more information relative to an affected child as the degree of dominance increases. This is evidenced in Table IV by the general trend for the ratio of affected ELOD to unaffected ELOD to decrease with increasing degree of dominance of the model. Even under the most dominant model, however, an affected child provides over twice the information of an unaffected child.

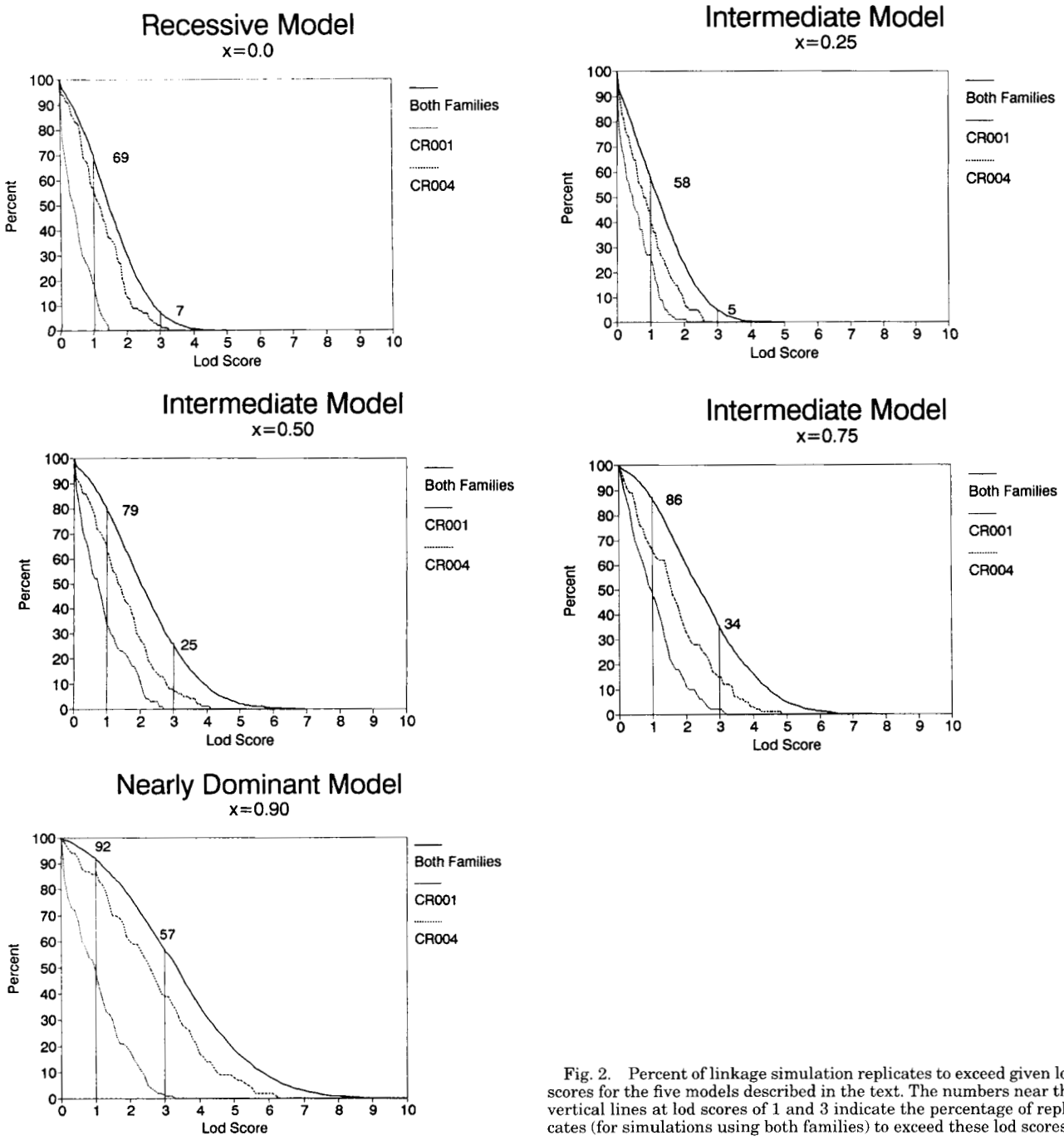

Fig. 2. Percent of linkage simulation replicates to exceed given lod scores for the five models described in the text. The numbers near the vertical lines at lod scores of 1 and 3 indicate the percentage of replicates (for simulations using both families) to exceed these lod scores.

DISCUSSION

In this report we describe a set of Costa Rican families which are informative for linkage analysis for BP. The particular utility of this study sample derives from the availability for genotyping studies of a large number of individuals with BP-I illness who have been diagnosed using a rigorous approach, and who are members of a few extended pedigrees. There is no evidence that the proportion of individuals with a diagnosis of BP-I compared to those with other mood disorder diagnoses is higher in this sample than in others that have been described for linkage studies of BP. Nevertheless, as demonstrated by the results of simulation analyses the size of the families enables us to define the BP phenotype very narrowly (BP-I and SAD-M only) and still retain sufficient power to detect evidence suggestive of linkage under several models of disease transmission. It must be considered, however, that the BP-I classification, as with all other recognized psychiatric diagnoses, is entirely based on clinical grounds, and has no independent validity. It is possible that identifying linkage may require even more narrow definitions of

TABLE III. Percentage of Simulation Replicates to Meet or Exceed a T Value of 2.32 or 3.71*

	Model and T value									
	I		II		III		IV		V	
Family	2.32	3.71	2.32	3.71	2.32	3.71	2.32	3.71	2.32	3.71
1	7	1	16	5	20	4	25	10	25	6
4	8	3	21	8	21	5	33	16	43	19
B	9	1	36	8	32	8	49	22	52	23

* These correspond to a (1-sided) P of 0.01 or 0.0001, and are comparable to lod scores of 1.16 and 3.0, respectively. B = both families combined.

phenotype, for example by age of onset; future simulations will establish the power of our sample considering such subdivisions among the individuals with diagnoses of BP-I.

Several conclusions can be drawn from the linkage simulation results. First, in the case of a major locus without locus heterogeneity between the families, the linkage method is almost certain to detect evidence of the location of the disease gene, under all single locus models of inheritance. Second, there is a considerable amount of information even in the possible presence of heterogeneity, that is, if different disease susceptibility loci were present in CR001 and CR004; in CR004 alone a lod score of ≥ 1 would be expected in the majority of cases under all models.

The conservatism of the models that we have utilized can be contrasted with those employed in linkage analyses of BP in other large pedigrees, for example, in the initial evaluation of chromosome 11 markers in the Old Order Amish. There, a gene frequency for BP of 0.021 was used, leading to a population prevalence of 3.63%. All of the models used in our analysis lead to a population prevalence of 1.5% for BP-I, the most narrowly defined BP phenotype. In the Amish analyses only 2.6% of all cases in the population would be phenocopies; with the parameters employed in our analyses, 67% of all cases would not be carriers of the disease gene. Nevertheless, the simulations indicate that there is a high probability of obtaining evidence suggestive of linkage, even using these conservative parameter estimates. The simulations demonstrate, however, that only in a relatively small percentage of marker analyses would we detect usually accepted thresholds of evidence for linkage using these assumptions. The wide range of possible lod scores is due to the incomplete informativeness in any given family of most highly polymorphic markers; given the occurrence of several large sibships in these families, for those markers in which a parent is homozygous, information is lost for several meioses. The potential availability of multiple markers eliminates the need to maximize the information obtainable from linkage data by designating individuals with less severe phenotypes as "affected."

In the simulation analyses, the dominant model yields the most power. This result was expected as the families were chosen in large part due to multiple cases being ascertained across three or more generations. However, the degree of information contained in the screening samples is sufficient to permit detection of "candidate" regions even with models that involve vastly different assumptions (e.g., pure recessive). The power of the sample across the spectrum of models is important in that it is possible for the mode of transmission to differ between families or populations when susceptibility is due to a single major locus but especially so when heterogeneity exists. The power to detect interesting lod scores in the two families, when considered separately, provides partial protection against the poor performance of linkage analysis under conditions of extensive heterogeneity. The fact that the families are drawn from a relatively homogenous population with extensively documented genealogy provides further reassurance.

TABLE IV. Comparison of the Linkage Information Provided by Affected and Unaffected Individuals Under Five Models of Genetic Transmission*

Model	Affection status	Type 1	Type 2	Type 3	Weighted ELOD	Ratio
1	Affected	0.201	0	0.380	0.220	72.50
	Unaffected	0.131	0	0.054	0.003	
2	Affected	0.063	0.17	0.16	0.171	52.64
	Unaffected	0.115	0.003	0.056	0.003	
3	Affected	0.019	0.191	0.101	0.190	12.57
	Unaffected	0.090	0.015	0.066	0.015	
4	Affected	0.004	0.197	0.081	0.198	4.18
	Unaffected	0.051	0.047	0.096	0.047	
5	Affected	0.001	0.201	0.077	0.201	2.30
	Unaffected	0.017	0.087	0.140	0.087	

* Types 1, 2, and 3 refer to the 3 mating types described in the text. The weighted ELOD was obtained by considering the frequency of each mating type as described in the text.

As discussed above, linkage simulation analyses provide information on expectable lod scores, for a given data set, under a range of models of transmission. In the analyses presented in this paper several parameters were held constant. It is possible, however, that in some circumstances models based on very different assumptions will yield similar linkage results. For complex diseases, such as BP, in which there is relatively little available information to determine appropriate parameters, it is likely that different investigators will employ different models. Simulations are also useful in comparing the assumptions on which different models rest. Under the model used to analyze the Amish pedigree, the ratio of affected ELOD to unaffected ELOD (again, weighted over the three informative mating types) was 1.99. It is noteworthy that this ratio is less than that for any of the models used in our analysis of the Costa Rican pedigrees. Because a greater proportion of the information in these models derives from affected individuals, there is less likelihood that substantial changes in lod scores will result from new onsets of illness over the course of the study. The difference in these ratios between the models adopted in this study and those used in analyzing the Amish data are particularly striking given our conservative strategy of setting the rate of phenocopies in the population at 67% compared to 2.6% for the Amish model.

Considering that extensive heterogeneity cannot be excluded, and that data in support of the use of specific genetic parameters are limited, it is important to consider methods of genetic mapping which complement lod score analysis. Although the models that we have adopted cover the spectrum of possible modes of single gene transmission, it is conceivable that we would fail to detect linkage because of inaccurate estimation of disease gene prevalence or the rate of phenocopies. In this eventuality, the APM is a useful adjunct, as the results obtained are not dependent on correct parameter estimates. Although the APM is generally not as powerful as the lod score method, in these pedigrees, the probability of detecting a possible locus (e.g., T score ≥ 1) is more than 50% in the combined data set for all models except the fully recessive one. In addition, as evidenced by the simulation results under models I and II, under some conditions the APM may provide more power to detect a BP locus than the lod score method. It should be noted, however, that the assumptions about the distribution of the test statistic used in the APM depends on the analysis of large numbers of families; the distribution of the statistic in small numbers of families remains to be assessed.

Neither the lod score method nor APM are likely to detect linkage under conditions of extensive locus heterogeneity within families. Under such conditions, association methods, using samples of patients, may be more powerful. In parallel with the pedigree investigations described in this paper, we have collected a sample of apparently unrelated BP-I individuals from the Costa Rican population. As discussed elsewhere [Escamilla et al., submitted] this sample is suitable for searching for genome regions where genetic association across loci (linkage disequilibrium) is conserved based on identity by descent from common ancestors who carried the disease mutation. As has been previously demonstrated, these regions can be detected with great efficiency even under circumstances of etiologic heterogeneity [Houwen et al.,1994].

ACKNOWLEDGMENTS

This work was supported by grants from the NIH (MH49499, MH48695, and MH00916, to NBF, and MH47563, S. Barondes, PI) and a Young Investigator Award to NBF from the National Alliance for Research in Schizophrenia and Depression. We appreciate the support of the National Psychiatric Hospital, Costa Rica, the Social Security System of Costa Rica, and especially the members of families CR001, CR002, CR003, and CR004. We thank Carmen Araya for clerical and organizational assistance and Maria-Luisa Cruz for clinical interviewing. We thank S. Barondes for helpful comments on the manuscript and for his longstanding support of this project.

REFERENCES

Adis G (1992): "Disordenes Mentales en Costa Rica: Observaciones Epidemiologicas." San Jose, Costa Rica: Editorial Nacional de Salud y Seguridad Social.

American Psychiatric Association (1987): "Diagnostic and Statistical Manual of Mental Disorders," Third Edition, Revised. Washington, DC: American Psychiatric Association Press.

Angrist M, Kauffman E, Slaugenhaupt SA, Matise TC, Puffenberger EG, Washington SS, Lipson A, Cass DT, Eyna TR, Weeks DE, Sieber W, Chakravarti A (1993): A gene for Hirschsprung disease (megacolon) in the pericentromeric region of human chromosome 10. Nat Genet 4:351–356.

Baron M, Freimer N, Risch N, Lerer B, Alexander JR, Straub RE, Asokan S, Das K, Peterson A, Amos J (1993): Diminished support for linkage between manic depressive illness and x-chromosome markers in three Israeli pedigrees. Nat Genet 3:49–55.

Bertelsen A, Harvald B, Hauge M (1977): A Danish twin study of manic-depressive disorders. Br J Psychiatry 130:330–351.

Cannon-Albright LA, Goldgar DE, Meyer LJ, Lewis CM, Anderson DE, Fountain JW, Hegi ME, Wiseman RW, Petty EM, Bale AE (1992): Assignment of a locus for familial melanoma, MLM, to chromosome 9p13-p22. Science 258:1148–1152.

Egeland JA, Gerhard DS, Pauls DL et al. (1987): Bipolar affective disorders linked to DNA markers on chromosome 11. Nature 325: 783–755.

Endicott J, Spitzer RL (1979): Use of research diagnostic criteria and the schedule for affective disorders and schizophrenia. Am J Psychiatry 137:52–59.

Escamilla M, Spesny M, Reus VI, Gallegos A, Meza L, Sandkuijl LA, Fournier E, Leon PA, Smith LB, Freimer NB: The use of linkage disequilibrium approaches to map genes for bipolar disorder in the Costa Rican population. 67:244–253.

Freimer NB, Reus VI, Escamilla MA, McInnes LA, Spesny M, Leon P, Service SK, Smith LB, Silva S, Ruja E, Gallegos A, Meza L, Fournier E, Bahorloo S, Blankenship K, Tyler DJ, Batki S, Vinogradov S, Weissenbach J, Barondes S, Sandkuijl LA (1996): Genetic mapping, using haplotype, association and linkage methods, suggests a locus for severe bipolar disorder (BPI) at 18q22–18q23. Nat Genet 12:436–441.

Gallegos A (1988): Estructuracion de un plan nacional de salud mental. Revista Medica de Costa Rica 55:503.

Goodwin FK, Jamison KR (1990): "Manic-Depressive Illness." New York: Oxford University Press, pp 373–401.

Gruis NA, Sandkuijl LA, Weber JL, van der Zee A, Borgstein A-M, Bergman W, Frants RR (1993): Linkage analysis in Dutch familial atypical multiple mole-melanoma (FAMMM) syndrome families: Effect of naevus count. Melanoma Res 3:271–277.

Heutink P, van der Mey AG, Sandkuijl LA, van Gils AP, Bardoel A, Breedveld GJ, van Vliet M, van Ommen GJ, Cornelisse CJ, Oostra BA (192): A gene subject to genomic imprinting and responsible for

hereditary paragangliomas maps to chromosome 11q23-qter. Hum Molec Genet 1:7–10.

Houwen RHJ, Baharloo S, Blankenship K, Raeymaekers P, Juyn J, Sandkuijl LA, Freimer NB (1994): Genomic screening by searching for shared segments: Mapping a gene for benign recurrent intrahepatic cholestasis. Nat Genet 8:380–386.

Kelsoe JR, Ginns EI, Egeland JA, Gerhard DS et al. (1989): Re-evaluation of the linkage relationship between chromosome 11p loci and the gene for bipolar affective disorder in the Old Order Amish. Nature 342:238–243.

Leckman J, Sholomskas D, Thompson W, Belanger A et al. (1982): Best estimate of lifetime psychiatric diagnosis: A methodological study. Arch Gen Psychiatry 39:879–883.

Leon PE, Raventos H, Lynch E, Morrow J, King MC (1992): The gene for an inherited form of deafness maps to chromosome 5q31. Proc Natl Acad Sci (USA) 89:5181–5184.

Nurnberger JI, Blehar MC, Kaufmann CA, York-Cooler C et al. (1994): Diagnostic interview for genetic studies. Arch Gen Psychiatry 51:849–859.

Pauls DL, Morton LA, Egeland JA (1992): Risks of affective illness among first-degree relatives of bipolar I old-order Amish probands. Arch General Psychiatry 49:703–708.

Pauls DL, Pakstis AJ, Kurlan R, Kidd KK, Leckman JF, Cohen DJ, Kidd JR, Como P, Sparkes R (1990): Segregation and linkage analysis of Tourette's syndrome and related disorders. J Am Acad Child Adolesc Psychiatry 29:195–203.

Saborio M (1992): Experience in providing genetic services in Costa Rica. *Birth Defects Original Article Series* 28 (3):96–102.

Spence MA, Ameli H, Sadovnick AD, Remick RA, Bailey-Wilson JA, Flodman P, Yee IML (1993): A single major locus is the best explanation for bipolar family data: results of complex segregation analysis. Am J Hum Genet 53:862.

Spitzer RL, Endicott J, Robins E (1978): Research diagnostic criteria: Rationale and reliability. Arch Gen Psychiatry 35:773–782.

Terwilliger JD, Ott J (1994): "Handbook of Human Genetic Linkage." Baltimore: The Johns Hopkins University Press.

Weeks D, Lange K (1988): The affected pedigree-member method of linkage analysis. Am J Hum Genet 42:315–326.